DOCUMENT RESUME

ED 477 344 EA 032 324

AUTHOR	Goldhaber, Dan
TITLE	What Might Go Wrong with the Accountability Measures of the "No Child Left Behind Act"?
PUB DATE	2002-02-13
NOTE	16p.; Paper prepared for the "Will No Child Truly Be Left Behind? The Challenges of Making This Law Work" Conference (Washington DC, February 13, 2002).
PUB TYPE	Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; *Academic Standards; *Accountability; Educational Improvement; Educational Legislation; Educational Objectives; Educational Policy; Elementary Secondary Education; Federal Regulation; *Government Role; Government School Relationship; *National Standards; Outcomes of Education; Politics of Education; School District Autonomy; State Standards
IDENTIFIERS	*No Child Left Behind Act 2001

ABSTRACT

This paper explores the potential pitfalls associated with the new federal accountability role precipitated by the passage of the No Child Left Behind Act. The paper presents worst-case scenarios with the assumption that it is worthwhile to consider the potential for unanticipated consequences so as to avoid problems before they occur. After providing a general overview of the new federal, state, and local accountability relationship, the paper focuses on how accountability systems might create unanticipated negative consequences. More specifically, the paper discusses ways of misrepresenting educational realities; the problems of teaching to the test; shaping the pool of students to be tested; how the definition of a school may be manipulated to meet standards; adjusting state standards downward; tallying methods used for measuring progress; and possible checks on manipulating the system. The paper concludes that it would be unfortunate if manipulation and abuse of the law actually occurred because it would reduce the likelihood that the goals of the legislation would be realized, and it would undermine, in the eyes of the public, the notion that standards and accountability can be used to improve education. (Contains 34 references.) (WFA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 477 344

What Might Go Wrong with the Accountability Measures of the "No Child Left Behind Act?"

Dan Goldhaber

The Urban Institute

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. GOLDHABER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Prepared for

Will No Child Truly Be Left Behind?
The Challenges of Making This Law Work

A Conference Sponsored by the Thomas B. Fordham Foundation

February 13, 2002

EA 032 324

BEST COPY AVAILABLE

What Might Go Wrong with the Accountability Measures of the "No Child Left Behind Act?"

Dan Goldhaber
The Urban Institute

On January 8, 2002, President Bush signed the reauthorization of the Elementary and Secondary Education Act (also referred to as the "No Child Left Behind Act"). In many ways the passage of this legislation marked a significantly more prominent federal role in education. This is especially true with regard to the accountability provisions, which suggest that the federal government will, for the first time, penalize schools that fail to achieve "adequate yearly progress," as defined by student performance on standardized tests. Rewards and sanctions are, of course, designed to lead to better student outcomes, but incentives that are not properly structured may result in policies and behaviors that are not universally beneficial. In this memorandum, I explore the potential pitfalls associated with this new federal accountability role. In doing so I am not arguing that these worst case scenarios described below are likely, only that it is well worth the time to consider the potential for unanticipated negative consequences so as to try to avoid pitfalls before they occur.

There are, of course, many potential unanticipated negative consequences associated with any accountability system, be it at the local, state, or national level. After providing a general overview of the new federal, state, and local accountability relationship, I will focus on how accountability systems may create unanticipated negative consequences. The hope is that by pointing out the possible pitfalls associated with a federal role in accountability, these pitfalls may be avoided.

Overview of the New Federal Role

The centerpiece of the new federal role in accountability is the requirement that states administer *high-quality* annual academic assessment tests in reading and math for every child in grades three through eight by the 2005-06 school year. (In 2007-08 schools

will also be required to administer annual tests in science.)[1] These assessments must be *aligned* with standards, *consistent* with nationally *recognized* professional and technical standards, be *used in a valid and reliable manner*, and *test higher order thinking skills* using multiple measures.

Each state is required to create a system of rewards and sanctions based on whether students from a number of different sub-groups make adequate yearly progress (AYP) towards the state's *proficient* level of academic achievement.[2] AYP must be defined so that in each state all students in each group meet or exceed the state's proficient level of academic achievement "not later than twelve years after the end of the 2001-2002 school year" (2013-14).[3] Schools that fail to demonstrate AYP for two consecutive years are required to provide students with additional public school choices. If schools fail to improve after a third year, parents of students in those schools may use a portion of the school's Title I aid to purchase supplemental educational services, including private tutoring. Schools failing to improve for five consecutive years may be subject to reconstitution.[4] The legislation also requires states to participate in the National Assessment of Educational Progress (NAEP) in reading and math, which means a sample of students from the state will take this national proficiency test in grades 4 and 8. Student performance on the NAEP will be used to *verify* reported performance on the assessments used in each state.

While few argue against "appropriate" accountability measures, debate arises in regards to what is appropriate, and, in the case of the reauthorization of ESEA, the devil is very much in the details, many of which are sketchy and left open for negotiation between states and the Department of Education. For example, the question of what

---

[1] This is not by any means a comprehensive portrait of the accountability portion of the legislation. For instance, the legislation also specifies intermediate goals, including statewide annual measurable objectives to meet this long-term objective. Public Law 107-110, Title I, Part A, Subpart 1, Section 1111(b)(2)(H).

[2] These subgroups include racial, ethnic, and economic groups, as well as students with disabilities and those with Limited English Proficiency.

[3] Twelve years from the end of the 2001-2002 school year would be beyond a second term of the Bush administration so policy priorities may change before this deadline. The legislation specifies intermediate goals for meeting this objective. These include each state establishing "statewide annual measurable objectives" that indicate a "single minimum percentage of students who are required to meet or exceed the proficient level on the academic assessments." These minimum percentages apply separately to each subgroup of students and not all subgroups must make adequate yearly progress each year. See Public Law 107-110, Title I, Part A, Subpart 1, section 1111(b)(2)(F) through (I).

[4] Reconstitution of a school refers to the re-evaluation of all personnel staffing positions at that school.

constitutes adequate yearly progress received a great deal of attention.[5] AYP along with the other italicized words and phrases in the preceding two paragraphs (e.g. "high-quality," "proficient," "verify") are somewhat vague and certainly open to debate. What constitutes a "high-quality" assessment? How do we know whether assessments are aligned and consistent with recognized professional standards? What precisely does it mean to use an assessment in a valid and reliable manner? What is academic proficiency? What constitutes verification of a state's assessment results? Can the NAEP results be used to do this?[6]

These are certainly all important questions that create considerable disagreement among policymakers and academics. The vagueness associated with many of the provisions in the ESEA may lead to educational progress by allowing for wise policymaking as states and the federal government work together to craft policies that best fit specific local contexts. But it is also possible that this vagueness will work to the detriment of education as states, localities, and schools game accountability systems so as to best demonstrate that adequate yearly progress is being achieved.

Ways to Misrepresent Educational Realities

In recent years, standards-based reform and accountability has become a central component of school reform initiatives in most states. Virtually all states now have developed academic standards that students are expected to meet and tests to judge school and student performance against those standards.[7] In theory this guarantees that state officials, as well as the public at large, know how much students in the state are learning. But, there are a number of ways for school districts, schools, and teachers to make it appear that their students are learning more than they actually are. The most direct is outright cheating on state assessments, a method that has been used in the past on a number of occasions.[8] Other subtle (and legal) methods may also be used to either

---

[5] This was in part because of a study (Kane, Staiger, and Geppert, 2001) showing that an overwhelming number of elementary schools in North Carolina, a state widely regarded as having a sophisticated accountability system that has resulted in improved student outcomes (Grissmer and Flanagan, 1998), would have been judged as failing based on some of the originally-proposed AYP standards.

[6] This question is addressed elsewhere in this report.

[7] A number of states also are attaching "high-stakes" to these exams (Education Week, 2002).

[8] For example, in May 2001 a Maryland middle school suspended seven employees for suspected cheating on state exams (Slobogin, 2001). In 1999, a cheating scandal affected teachers in schools across New York

achieve or show educational gains that are not as large as they may appear on first blush. These fall under several general headings: *strategic allocation of teacher effort; the shaping of the tested pool; the makeup of a school*; *"adjustments" of states' standards*; and *tallying methods used to measure progress*.

Strategic Allocation of Teacher Effort

Probably the most common critique of accountability systems that are based on student performance on standardized tests is that they create incentives for teachers to focus their efforts on the assessments for which they (or their schools) will be held accountable. In common parlance, they will "teach to the test." Though it is common to refer to this practice with a negative connotation attached, the practice is clearly not in and of itself a bad thing.[9] Teaching to a "good" test would be quite beneficial were it to encourage teachers to focus on class material that is educationally beneficial to their students. Thus, the accompanied implicit assumption is that teaching to a test causes teachers to focus on topics deemed to be educationally unimportant for students in the long-run.[10] The curriculum itself is often said to become "narrowed" so as to focus only on tested material. For example, teachers may focus their efforts on tested subjects, such as math and English, at the expense of subjects that are not tested, such as science, a subject that is not required to be tested until 2007-08. Teachers may also spend their time simply teaching test taking skills (Education Week, 2001; Koretz et al., 1998; Schrag, 2000). Some research does suggest that accountability systems have led some teachers to incorporate standardized test content and test-taking skills into the curriculum at the expense of other material judged by many to be more educationally important (Education Week, 2001; Linn, 2000).

Another way teachers might strategically allocate their efforts is by focusing on only certain types of students (Elmore et al., 1996; Heubert et al., 1998). The new ESEA legislation requires the use of a system, already in place in many states, whereby schools'

City, while in 2000 Michigan elementary and middle schools were suspected of cheating on state exams (Hoff, 1999; Keller, 2001).

[9] See, for example, Yeh, 2001.

[10] Emerging research on states with high-stakes testing regimes, such as Texas and North Carolina, suggests that states' accountability systems are having positive effects on students' achievement (Grissmer

performance is judged based on the percentage of students who reach established benchmarks for proficiency.[11] Under such a system, it is the *pass rate* that matters for school performance, so schools have an explicit incentive to push as many students as possible beyond the point where they are judged to be proficient. This means that schools do not get credit for learning by students who are already above the proficiency level, nor do they get credit for learning by students who fail to jump the bar. Thus the system encourages a focus on those students who are just below the benchmark. Students far below the benchmark may be seen by teachers as "lost causes," and therefore not a good place to focus efforts. Research on the accountability system employed in Kentucky lends credence to this concern. It suggests that teachers have focused efforts on average or higher-achieving students to the detriment of lower-achieving students.

Shaping of the Tested Pool

One of the best ways for schools to influence accountability results is to shape which groups of students take a test. In general, the higher the percentage of students who sit for an exam, the lower the average score on that exam (or alternatively, the lower the pass rate on the exam). This is because the highest achieving students are the ones who are most likely to sit for exams on any given day. This is the reason many states require a certain percentage of students to be tested for a school to qualify for exemplary accountability ratings, and why some states explicitly factor in attendance on the day of the test when judging a school's performance (Education Week, 2001). There are, however, a number of ways that states can strategically manipulate the tested pool without showing lower attendance rates.

In the past, one way schools could manipulate their scores was by placing students into non-tested categories, such as Special Education and English Language Learners (ELL).[12] Such categories are sometimes exempt from testing and have mainly

---

et al. 1998). The evidence connecting accountability systems to improved student performance is not, however, conclusive (Haney, 2001).

[11] Texas, for example, has an Accountability Rating System, which is based on the percentage of students in the total population and certain subgroups who reach a established benchmarks on the state assessment (the TAAS). In order for a school to receive a "recognized" rating in Texas, at least 80% of the total students and each student subgroup must pass each TAAS subject test.

[12] Research on the classification of students into special education categories suggests that teacher referrals for special education services are many times improperly based on student characteristics such as race,

been exempt from counting toward schools' accountability ratings. The 2001 reauthorization of ESEA explicitly requires states to assess the achievement of students with disabilities and limited English proficiency and it requires *all* students to reach proficient levels after 12 years. This may lead to a greater focus on disabled and ELL students. One wonders, however, how exactly those provisions will work. The explicit requirement that these special classes of students be included in the accountability system goes beyond the provisions of the 1994 law that required states' standards and assessments apply to *all* students, including special education students and ELLs. Many states sought and obtained waivers from these requirements or ignored them altogether (Taylor, 2002). Even if there is strict enforcement of the 2002 law, one still might argue that incentives exist for classification of students into these special categories since students with special needs are sometimes provided with testing accommodations.

Another way that schools may influence their testing pools is through promotion and retention policies. The new emphasis on accountability is likely to encourage schools to adopt even more stringent promotion and retention policies to ensure that students are not promoted to grades where they will perform poorly on state assessments and hurt the performance of the school. Schools, for instance, may be less likely to promote students with weaker academic skills into 3rd grade, which is the first grade with required testing.[13] This is not necessarily a negative consequence of the outcome since the jury is still out on the net impact of retention on students' ultimate outcomes.[14] The research consensus, however, is that retention increases the probability of students dropping out of high school (Holmes, 1989; Grissom and Shepard 1989). Haney (2001), for instance, finds that when an exit exam in Texas was first implemented, dropout rates increased substantially, especially for African-American and Hispanic students.

---

gender, and socio-economic status, rather than on a student's actual need for special services (Ortiz, 1992; Singhal, 1999; Artiles, 1994). There is little evidence on the factors influencing the classification of students into ELL status.

[13] Alternatively, they may hustle students with strong academic skills into the 3rd grade.

[14] Far more studies argue against retention than for it (Holmes, 1989), though some studies show positive academic benefits (Kerzner, 1982; Pierson and Connell, 1992; Karweit, 1999; Eide and Showalter, 2000).

The Makeup of a School

Up to this point, I have implicitly treated what constitutes a school as a given and focused on the shaping of the pool of students within schools. There are, however, some interesting ways in which some school districts or states might manipulate the definition of a school so as to make it appear that the "school" is making AYP. For example, school systems could define "schools" in such a way that they consist of specific grades or classrooms within a single building. School systems could also classify multiple distinct "school" buildings into what would be considered by states as "single" schools.[15] Thus, local school systems could, through aggregation and reclassification of "schools," have high-achieving students offset the poor performance of lower-achievers.

One can make essentially the same case for the drawing of school district boundaries. Through educational gerrymandering neighborhoods could, for instance, be carved up so students are grouped together to maximize the probability that the largest number of schools demonstrate AYP.[16] Virginia's accountability system provides an excellent example of the potential for this type of manipulation. The unit of analysis in the Virginia accountability system is the *school*, not the *students* in the school. Thus, schools in the state may move in and out of accredited status simply based on the catchment areas of those schools. In other words, an accredited school one year could be unaccredited the next because different (lower-achieving) students are redistricted into a particular school building and this clearly is not related to the performance of personnel within the school.

"Adjustments" of States' Standards: A Race to the Bottom?

The re-authorized ESEA mandates that all states establish proficiency levels that all students in the state meet or exceed by 2013-14, but, as I mention above, it is not specific about what constitutes proficiency or how this should be measured. The language in the legislation mandates that state assessments conform to "*recognized* professional and technical standards," but an examination of various state assessments

[15] States receive student achievement information based on school codes. There is nothing that precludes states from allowing districts, for example, to specify two "school" buildings from opposite ends of a county as having the same code. From a state's perspective, this would then *de facto* be the same school.

used today suggests that there are in fact no universally held views about what constitutes "good" standards.[17] In fact, various groups rate states' standards quite differently in some cases. For example, under *Education Week*'s Standards and Accountability ratings, Kentucky receives an A- but the Fordham Foundation rates Kentucky as having "Trouble Ahead," meaning strong accountability attached to bad standards. Furthermore, there exists today a surprising amount of variation among states in how they rate the performance of their students in Title I (lower income) schools (U.S. Department of Education, 2001). For example, in Georgia 59 percent of Title I schools were identified as being in need of improvement while Tennessee identified only 2 percent of its Title I schools.[18] Were states to set the bar low enough, 100 percent of their students could be judged as proficient today.

Tallying Methods Used for Measuring Progress

The reauthorization of the ESEA is also silent on the precise methodology that states should use to measure or tally progress toward meeting the goals outlined in the legislation. The specific attributes of accountability systems differ significantly between states. For instance, among the states that use tests, there is variation in the type of exam used to measure student achievement. Some use assessments developed by the state (e.g., TAAS), while others use norm-referenced tests (NRT) such as the Stanford-9. Still others employ criterion-referenced tests (CRT), such as the Terra Nova. Many states use a combination of these options. States may use different tests from one year to the next, and these may not be designed to be directly comparable from year to year. The reason is that NRTs show how students in a particular grade compare relative to other students at a particular grade level, while CRTs show the extent to which students have mastered particular skills. It is possible for students in a particular state to improve their performance on CRTs while they perform less well on NRTs (or vice versa), particularly if states adopt different standards. This combination would reflect students who are gaining proficiency on their state's standards but who are not performing as well *relative*

---

[16] This would not work indefinitely because, holding the true achievement levels of students constant, there are only so many ways that high- and low-achieving students can be grouped to show AYP over time.
[17] Public Law 107-110, Title I, Part A, Subpart I, Section 1111(b)(3)(C)(iii).
[18] U.S. Department of Education, 2001.

to other students (often nationally) on the items on the NRT (which may not be closely aligned with their particular state's standards). The result of using very different types of assessments is that it would be necessary to use some secondary method to determine academic growth from year to year and thus comply with the AYP mandate. This, of course, is not a trivial or uncontroversial task.

There are also major differences in the tallying methodologies used to assess school performance. Today states use a variety of accountability standards, such as the average scores by grade level, the percentage of students who reach established benchmarks, changes over time in these measures, and various "value-added measures" such as the school-level average of gains for individual students.[19] Some are far better than others at identifying the actual contributions of teachers and schools.[20] But regardless of the system employed, it is common to observe the so called "saw-tooth effect" — the finding that that test scores increase substantially during the initial years of a test's administration due simply to increased familiarity with the assessments, and then level off (Heubert, 1998; Koretz, 1988; Linn, 2000; Schrag, 2000).

If test scores do increase substantially during the initial years of their administration and then level off, states might introduce new assessments once they have reached the leveling off point. States may also simply change the rules of the tallying system. In Virginia, for instance, starting in 2001, the state changed the methodology used to determine schools' performance on the Standards of Learning (SOL) test, the state's assessment. The difference between the scores under the old and new methodology is that the new scores account for the performance of students who had previously failed to reach proficiency levels but had been through a remediation program and retaken the test. These students, however, are only accounted for in the numerator. This adjustment to the accountability system in Virginia has created the strange situation where, at least in theory, schools can have adjusted SOL pass rates of over 100 percent even if the majority of students at a particular grade level were not judged to be

---

[19] Other value-added measures include comparing differences between actual and regression-generated predicted scores.

[20] For instance, in my opinion, it is necessary to use a value-added methodology and account for family, student, and background factors to effectively isolate the contributions of schools and teachers. Additionally, most standardized achievement tests are designed to provide relative scores and they may be inadequate at measuring whether students have mastered particular standards (Popham, 2002).

proficient. This new method of calculating pass rates also makes it appear as if the state is making greater progress towards the goal of all students in the state achieving academic proficiency.

There may well be valid reasons for Virginia altering their method to assess schools, however, it illustrates the point that such systems can be manipulated simply for the sake of changing perceived progress. The bottom line is that accountability systems may be gamed to show student achievement gains. This is possible because states have the flexibility to set their own standards, administer their own tests, and craft systems to judge student performance. Thus, one could imagine a worst case scenario where the pressure, political and otherwise, to show that students are making academic gains could create a race to the bottom in terms of standards and accountability systems.

Conclusions: Checks on the Gaming of the System?

What is to prevent states from setting low standards or the manipulating of the system of the sort described above. In theory the highly regarded national proficiency test administered in grades 4 and 8—the National Assessment of Educational Progress (NAEP) can be used to verify the reported state gains in academic proficiency. Serious manipulation of a state's system might be detected by discrepancies between state reports of students' AYP (based on state assessments) and their performance on NAEP. But, for a variety of reasons, there is considerable doubt as to whether NAEP is up to this task. One can easily imagine situations where states truly show remarkable student gains on the state assessment, but have their NAEP scores remain flat. This can occur, for instance, if a state opts to adopt standards that are not well-aligned with what is tested on the NAEP. Recent studies, in fact, have found a number of cases where states with large improvements in state test scores experienced little improvement on the NAEP (Klein et al., 2000; Koretz et al., 1998).

Discrepancies between state assessment and NAEP results would, of course, not preclude state officials from making the argument that their students are in fact gaining academically. Disputes over differences between NAEP and state assessment results will no doubt create a windfall for statisticians and testing experts in the business of equating different tests — this may be particularly difficult if many students opt out of taking the

NAEP test, as they are allowed to do. The truth about student achievement will be out there, but policymakers and much of the public likely will not know what to make of the arcane statistical arguments.

A second potential check on states gaming the system is the requirement that states' educational plans be approved by the Department of Education. But, the legislation also limits the Secretary's authority by explicitly stating that the Secretary "shall not have the authority to require a State, as a condition of approval of the State plan, to include in, or delete from, such plan one or more specific elements of the State's academic content standards or to use specific academic assessment instruments or items."[21] Furthermore, unlike the provisions in an earlier proposed version of the legislation, the Secretary does not have the authority to withhold educational funding from states that are not seen to be making AYP based on the NAEP. Thus, in some respects, the Secretary of Education wields a relatively soft stick. The bottom line is that political realities will likely place some major constraints on the ability of the Secretary to influence states' educational plans. As Toch (2001) notes, there has been far less than full adoption of the testing requirements that were put in place in the 1994 reauthorization of the ESEA.

The law takes what appears to be a firm stand that *all* students be proficient in 12 years, but this is an eternity in political terms. In the meanwhile, there exists a great deal of room to make it look like real progress is being made while the reality is otherwise. It would be truly unfortunate if manipulation of the sort described above actually occurred because it would reduce the likelihood that the goals of the legislation are realized and likely serve to undermine, in the eyes of the public, the notion that standards and accountability systems can be used as a means of improving education.

---

[21] Public Law 107-110, Title I, Part A, Subpart 1, Section 1111(e)(1)(F).

References

Artiles, Alfredo (1994). "Overrepresentation of Minority Students in Special Education: A Continuing Debate." *The Journal of Special Education*. 27(4).

Bishop, John H (1989). "Is the Test Score Decline Responsible for the Productivity Growth Decline?" *The American Economic Review*. Vol 79 (1).

Carnoy, Martin, Susanna Loeb, and Tiffany Smith (2000). "Do Higher State Test Scores in Texas Make for Better High School Outcomes?" Stanford University: School of Education.

Education Week (2001). Quality Counts, 2001: A Better Balance—Standards, Tests, and the Tools to Succeed. *Education Week* 20(17).

Eide, E.R. and Showalter, M.H. (2000). "The Effect of Grade Retention on Educational and Labor Market Outcomes." Forthcoming in *Economics of Education Review*.

Elmore, Richard, Charles Abelmann, and Susan Fuhrman (1996). "The New Accountability in State Education Reform: From Process to Performance." In Ladd, Helen (Ed.). Holding Schools Accountable: Performance-Based Reform in Education. Washington, D.C.: The Brookings Institute.

Erickson, Ronald (1998). "Special Education in an Era of School Reform: Accountability, Standards, and Assessment." Federal Resource Center.

Grissmer, David and Ann Flanagan (1998). "Exploring Rapid Achievement Gains in North Carolina and Texas." Washington, D.C.: National Education Goals Panel.

Grissom, J.B., and Shepard, L.A. (1989), Repeating and dropping out of school. In L.A. Shepard and M.L. Smith (Eds.), *Flunking Grades: Research and Policies on Retention*. London: Falmer, pp. 34-63.

Haney, Walt (2001). "Revisiting the Myth of the Texas Miracle in Education: Lessons about Dropout Research and Dropout Prevention." Paper prepared for the Dropout Research: Accurate Counts and Positive Interventions Conference: Cambridge, MA.

Heubert, Jay and Robert Hauser (Eds) (1998). "High Stakes: Testing for Tracking, Promotion, and Graduation." Committee on Appropriate Test Use, National Research Council.

Hoff, David J. (1999). "N.Y.C. Probe Levels Test-Cheating Charges." *Education Week* (web edition). 15 December 1999.

Holmes, C.T. (1989). "Grade level retention effects: a meta-analysis of research Studies." In L.A. Shepard and M.L. Smith (Eds.), *Flunking Grades: Research and Policies on Retention*. London: Falmer, pp. 16-33.

Kane, Thomas, Douglas Staiger and Jeffrey Geppert (2001) "Assessing the Definition of 'Adequate Yearly Progress' in the House and Senate Education Bills." Unpublished manuscript, 2001. Available: http://www.dartmouth.edu/~dstaiger/WP.html

Karweit, N.L. (1999). "Grade Retention- Prevalence, Timing, and Effects." CRESPAR Report No. 33. Web address: http://scov.csos.jhu.edu/crespar/reports/report33chapt1.html.

Keller, Bess (2001). "Dozens of Mich. Schools Under Suspicion for Cheating." *Education Week* (web edition). June 20, 2001.

Kerzner, R.L. (1982). *The Effect of Retention on Achievement*. Union, NJ: Kean College of New Jersey.

Klein, Stephen P., Laura S. Hamilton, Daniel F. McCaffrey, Brian M. Stecher (2000). "What Do Test Scores in Texas Tell Us?" RAND.

Koretz, Daniel and Sheila Barron (1998). "The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)". RAND.

Linn, Robert (2000). "Assessments and Accountability." *Educational Researcher* 29:2.

Linton, Thomas (2000). "High Stakes Testing in Texas: An Analysis of the Impact of Including Special Education Students in the Texas Academic Excellence Indicator System." Paper presented at the Texas Assessment Conference, Austin, Texas.

National Center on Educational Outcomes (1999). "A Report on State Activities at the End of the Century." Available on the web: http://www.coled.umn.edu/NCEO/

Ortiz, Alba (1992). "Assessing Appropriate and Inappropriate Referral Systems for LEP Special Education Students." Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues: Focus on Evaluation and Measurement.

Pierson, L.H., and Connell, J.P. (1992). "Effect of Grade Retention on Self-system Processes, School Engagement, and Academic Performance." *Journal of Educational Psychology* 84, 300-307.

Popham, James W (2002). "Right Task Wrong Tool." *American School Board Journal*. 189(2), 18-22.

Robelen, Erik W. (2001). "States Sluggish on Execution of 1994 ESEA." *Education Week* (web edition). 28 November 2001.

Schrag, Peter (2000). "Too Good To Be True." *The American Prospect.*

Singhal, Rebecca (1999). "Revisiting Segregation: The Overrepresentation of Minority Students in Special Education." ERIC document.

Slobogin, Kathy (2001). "Cheating scandals test schools." 23 July 2001. www.cnn.com.

Taylor, William L., and Dianne M. Piche (2002). "Will New School Law Really Help." *USA Today*. 9 Jan 2002. A13.

Toch, Tomas (2001). "Bush's Big Test: The President's Education Bill is a Disaster in the Making. Here's How He Can Fix It." *The Washington Monthly*, November.

U.S. Department of Education (2001). "High Standards for All Students." January, 2001.

Wolf, Patrick and Bryan Hassel (2000). "Effectiveness and Accountability in Special Education, Part I: The Compliance Model." Thomas B. Fordham Foundation and Progressive Policy Institute.

Yeh, Stuart S. (2001). "Tests Worth Teaching To: Constructing State-Mandated Tests that Emphasize Critical Thinking." *Educational Researcher.* Vol. 30(9). pp. 12-17.

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: What Might Go Wrong with the Accountability Measures of the "No Child Left Behind Act?"	EA 032 324
Author(s): Dan Goldhaber	
Corporate Source:	Publication Date: Feb 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: [signature]	Printed Name/Position/Title: Dan Goldhaber / Research Assoc. Prof.
Organization/Address:	Telephone: 206-616-8793 / FAX:
	E-Mail Address: dgoldhab@u.washington.edu / Date: 6/10/03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Educational Management
5207 University of Oregon
Eugene OR 97403-5207
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2001)